DNA STATISTICAL ANALYSIS FOR STR DATA

A. SCOPE

Databases listing the allele frequencies for various nuclear DNA loci are available for several different major racial groups including African Americans, Caucasians, Hispanics, Asians and American Indians. The National Institute of Standards and Technology (NIST) population database will be used for data analyzed with the GlobalFiler STR amplification kit. When racial groups other than African Americans, Caucasians, Hispanics and Asians require statistical calculations, the FBI database may be used. Statistics are calculated using the "Popstats" program.

A.1 **FORMULAS**:

A.1.1 **HOMOZYGOUS**

If an individual is **homozygous** at a particular locus, the genotype frequency is calculated as $p^2 + p(1-p)\,\theta$, where ***p*** is the frequency of the allele and θ is a measure of population subdivision: θ =0.01 can be used for large populations (e.g., the general) and θ =0.03 for very small, isolated populations (e.g. American Indian and Amish).

Example: An individual has a genotype of 12, 12. The frequency of the 12 allele = 0.130.

genotype frequency = $(0.130)^2$+0.130(1-0.130)(0.01)
genotype frequency = 0.0169+0.130(0.87)(0.01)
genotype frequency = 0.0169+0.001131
genotype frequency = 0.018031

A.1.2 **HETEROZYGOUS**

If an individual is **heterozygous** at a particular locus, the genotype frequency is calculated as ***2pq***, where ***p*** and ***q*** are the frequencies of the respective alleles.

Example: An individual has a genotype of 9, 11. The frequency of the 9 allele = 0.145 and the frequency of the 11 allele = 0.214.

genotype frequency = 2 (0.145 X 0.214)
genotype frequency = 2 (0.03103)
genotype frequency = 0.06206

A.1.3 STOCHASTIC LEVEL SINGLE PEAKS

If an inclusionary forensic profile contains a single, stochastic level peak at a given locus, the genotype frequency is calculated as ***2p***, where ***p*** is the allele frequency of the respective allele. This allows for the consideration of allele drop-out due to stochastic effects at that locus.

Example: A forensic profile contains only one stochastic level 9 allele at a locus. The frequency of the 9 allele = 0.145.

genotype frequency = 2 (0.145)
genotype frequency = 0.29

A.1.4 OVERALL GENOTYPE

The joint probability across multiple loci can be calculated by multiplying their respective frequencies together. Single source, dominant, and minor DNA profiles are calculated as follows:

Example: An individual has the following types:

Locus 1: 12, 13 (allele frequency of 0.145 and 0.192)
Locus 2: 17, 25 (allele frequency of 0.002 and 0.046)
Locus 3: 17, 18 (allele frequency of 0.159 and 0.110)

Frequency = (Freq. of Locus 1) X (Freq. of Locus 2.) X (Freq. of Locus 3)
Frequency = (0.05568) X (0.000184) X (0.03498)
Frequency = 0.000000357
Probability = 1/genotype frequency
Probability = 1 in 2,801,120 (1/0.000000357 rounded)
Probability of observing the given genotype is reported as 1 in 2.80 million.

NOTE: Statistical information is not utilized regarding estimated frequencies for fragments detected at the amelogenin, Y-Indel and DYS391 loci for the GlobalFiler kit.

A.1.5 MINIMUM ALLELE FREQUENCY

Minimum allele frequency estimations are calculated for any allele that was not observed at least five times in the population database. The intent of this application is to set a lower limit for the frequency for such rare alleles and, consequently, produce a conservative allele frequency estimation that does not underestimate the allele's frequency of occurrence.

The approach utilized is a basic procedure described previously by the National Research Council, 1996.

The minimum allele frequency is calculated using the following expression:

$$P_{min} = 5/2n$$

Where **n** represents the sample size (individuals).

A.1.6 MIXTURE CALCULATIONS

A mixture calculation may be performed using the allele frequencies of each allele represented to determine the probability of selecting an unrelated individual in the population who could be a potential contributor to the mixture.

Frequency of occurrence calculations for an individual STR locus and DNA profiles with multiple STR loci when a mixture of DNA is present are calculated as follows:

The allele frequency for each allele in the mixed DNA profile for an individual locus is determined.

Individual fragment frequencies are summed then this value is squared:

$$(p_1 + p_2 + ...p_n)^2 = P_{LOCUS}$$

Where n equals the number of alleles present, and P_{LOCUS} is the probability of an unrelated individual being a contributor to that mixture at that locus.

The value of P_{LOCUS} for each locus will be multiplied together to generate the probability of an unrelated individual in the population being a contributor to that mixture for all loci examined.

$$P_{LOCUS1} \text{ X } P_{LOCUS2} \text{ ... X } P_{LOCUS(n)} = P_{MIX}$$

Where n equals the number of loci analyzed and P_{mix} is the probability of an unrelated individual being a contributor to the mixture.

Loci with any alleles below the stochastic threshold will not be used in the mixture statistical calculation.

Single source statistics may be applied to a minor profile where the minor profile can be clearly extracted from the mixture (Example 1). The mixture calculation can be used when calculating statistics for a minor profile where the minor contributor cannot be clearly distinguished (Example 2); however, the mixture calculation cannot be used for any loci where possible allelic drop out has occurred (Example 3).

Example 1

11 160 RFU
12 165 RFU
16 4000 RFU
18 4600 RFU

The 11, 12 alleles are clearly the minor profile. These alleles can be used in a single source calculation.

Example 2 (3130-1, 5 sec injection)

9 425 RFU
11 700 RFU
13 1000 RFU

The 9 allele is indicative of a minor allele in a two person mixture. All alleles are above the GlobalFiler stochastic threshold of 260 RFU; therefore, there is no indication of drop out and all alleles i.e. 9, 11, and 13 will be entered for this locus in the mixture statistical calculation.

Example 3 (3130-1, 5 sec injection)

9 250 RFU
11 750 RFU
13 950 RFU

The 9 allele is indicative of a minor profile. However, since the 9 allele has an RFU value below the GlobalFiler stochastic threshold of 260 RFU, drop out of a sister allele is possible. This locus will be left out of mixture statistics (CPI). Please note that in this instance, the "9" allele may be evaluated according to the "2p" statistical approach as described in section **A.1.3** of this document, and the assumed single source dominant be subject to an RMP calculation.

Example 4 (3500xL, 24 sec injection)

9 250 RFU
11 750 RFU
13 950 RFU

The 9 allele is indicative of a minor profile. However, since the 9 allele has an RFU value below the GlobalFiler stochastic threshold of 555 RFU, drop out of a sister allele is possible. This locus will be left out of mixture statistics (CPI). Please note that in this instance, the "9" allele may be evaluated according to the "2p" statistical approach as described in section **A.1.3** of this document, and the assumed single source dominant be subject to an RMP calculation.

A.1.7 SOURCE ATTRIBUTION

Identity/Source Attribution Calculation

Formula for not observing the profile in a sample of N unrelated individuals:

$p_x \leq 1-(1-\alpha)^{1/N}$

Using $\alpha = 0.001$ for 99.9% confidence; N = world population (Note: a population larger than the current world population is utilized to allow for less frequent updating of the threshold)

$p_x \leq 1-(1-0.001)^{1/8,000,000,000}$

$p_x \leq 1.25011E-13$ or the inverse $p_x \geq 1$ in 8,000,000,000,000

If the estimated frequency of the matching profile is rarer than 1 in 8 trillion individuals in each of the four populations examined, a source attribution statement can be made.

B. REFERENCES

B.1 DNA Technology in Forensic Sciences, National Research Council, National Academy Press, 1992.
B.2 The Evaluation of Forensic DNA Evidence, National Research Council, National Academy Press, 1996.
B.3 Forensic DNA Typing, John M. Butler, Academic Press 2001 and 2005.
B.4 SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, Approved 1/14/10.
B.5 Current population data as documented on the most up to date CODIS bulletin retained in the performance check binder.
B.6 NIST Population database web page: http://strbase.nist.gov/NISTpop.htm#Autosomal